

plot.design

青木繁伸

2020年3月17日

1 目的

カテゴリ変数のとる値ごとに、目的変数について指示した統計量を求め、一画面上に図示する。
Rにあるplot.designを書いた。

2 使用法

```
import sys
sys.path.append("statlib")
from plot import plot_design
plot_design(data, FUN=np.mean, verbose=True)
```

2.1 引数

data	データフレーム
FUN	使用する関数 (デフォルトは np.mean)
verbose	必要最小限のプリント出力をする

2.2 戻り値の名前

"vname"	最左列も含めた変数名
"value"	2 列目以降の変数のカテゴリ (二重リスト)
"result"	2 列目以降の変数のカテゴリごとの統計量 (二重リスト)
"FUN"	使用した関数

3 使用例

```
import numpy as np
import pandas as pd

x1 = np.repeat([1, 2, 3], 10)
x2 = np.repeat(["male", "female"], [5, 25]) # 右と同じ c(rep(1, 5),
```

```

    rep(2, 25))
y = [-0.37, -0.41, 0.98, -0.36, -0.32, 1.28, 0.95, 1.07, 1.13, -0.46,
-1.03, 0.78, 0.76, 0.54, 0.89, -1.05, -1.04, 1, -0.05, -0.8,
-0.17, 0.33, -0.19, 0.63, 0.04, -0.12, -2.03, 2.45, -0.44, -2.42]
data = pd.DataFrame({"y": y, "x1": x1, "x2": x2})

import sys
sys.path.append("statlib")
from plot import plot_design

a = plot_design(data)

```

function: mean

x1

1: 0.349

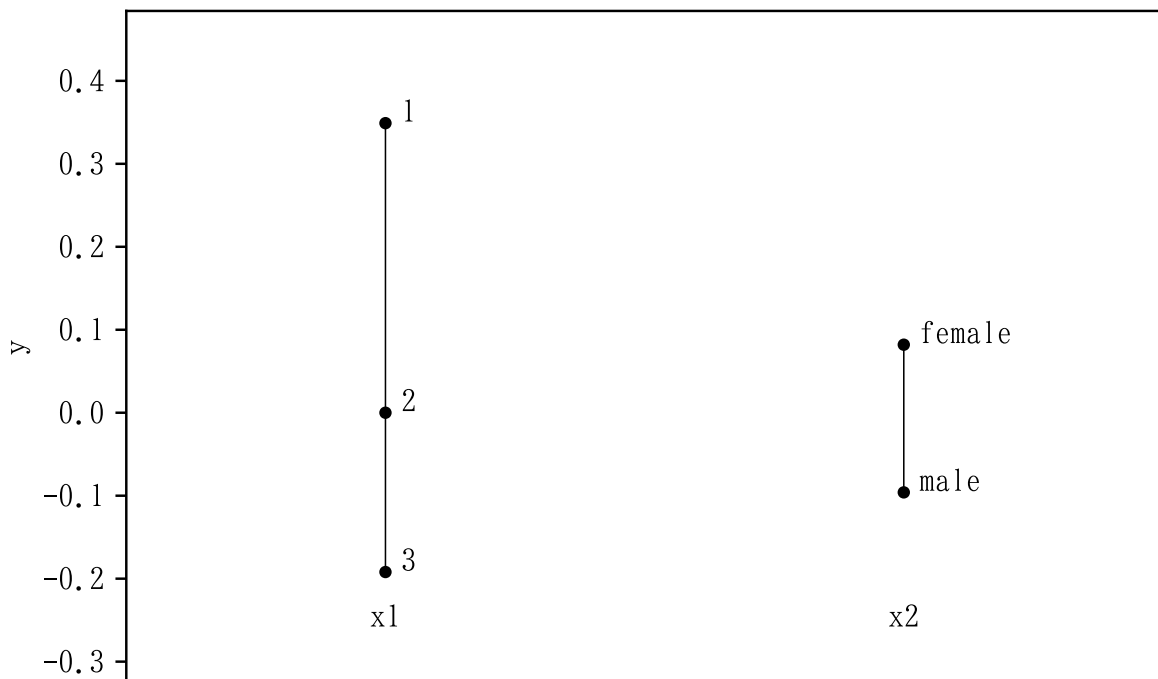
2: -1.1102e-17

3: -0.192

x2

female: 0.082

male: -0.096



```

np.random.seed(1239)
y = np.random.normal(50, 10, 200)
var1 = np.random.choice(["A", "B", "C", "D", "E"], 200)
var2 = np.random.choice(["hi", "med", "lo"], 200)
var3 = np.random.randint(10, 15, 200)
var4 = np.random.randint(1, 5, 200)

```

```
var5 = np.random.randint(1, 5, 200)
data = pd.DataFrame({"y": y, "var1": var1, "var2": var2, "var3": var3,
                    "var4": var4, "var5": var5})

a = plot_design(data, FUN=np.median)
```

function: median

var1

A: 53.599
B: 49.736
C: 50.338
D: 49.035
E: 47.978

var2

hi: 50.301
lo: 50.695
med: 47.204

var3

10: 47.539
11: 51.104
12: 49.799
13: 50.695
14: 49.337

var4

1: 47.037
2: 54.013
3: 50.228
4: 49.25

var5

1: 51.194
2: 47.161
3: 52.302
4: 50.16

